INFORMATICS INSTITUTE OF TECHNOLOGY

In Collaboration with

ROBERT GORDON UNIVERSITY ABERDEEN

Multimodal Music Information Classification System

A Project Proposal Document by

Mr. Sathila Samarasinghe

Mr. Thareen Withanage

Mr. Ishan Fernando

Mr. Rikzy Jezuli

Supervised by

Ms. Sachinthani Perera

Submitted in partial fulfillment of the requirements for the BSc (Hons) in Artificial Intelligence and Data Science degree at Robert Gordon University.

February 2023

Table of Contents

ROBERT GORDON UNIVERSITY ABERDEEN

INFORMATICS INSTITUTE OF TECHNOLOGY

List of Tables

List of Figures

List of Abbreviations

Acronym	Description
SVM	Support Vector Machine
CNN	Convolutional Neural Network
CALM	Categorizing And Learning Module
SATB	Soprano, Alto, Tenor, Base
CRNN	Convolutional Recurrent Neural Network
RNN	Recurrent Neural Network
ML	Machine Learning
KNN	K-Nearest Neighbor
Bi-LSTM	Bi-directional Long Short-Term Memory
DNN	Deep Neural Network
HMM	Hidden Markov Model
PLCA	Probabilistic Latent Component Analysis
AI	Artificial Intelligence
UI	User Interface
SSADM	Structured Systems Analysis and Design Method)
DSDM	Dynamic System Development Method
MMIC	Multimodal Music Information Classification
MMICS	Multimodal Music Information Classification System
ML	Machine Learning
API	Application Programming Interface
GUI	Graphical User Interface

1. CHAPTER 01: INTRODUCTION

1.1 Chapter Overview

In the modern era, people are finding a way to relax by listening to music on their smart devices. Commonly their approach would be downloading music tracks and categorizing them according to their choice into playlists. The primary domain of this research is to get rid of this traditional way of listening to music. Educating and classifying the information of music is the main scope of this project. Here the project mainly aims to help people to classify music into genres. Moreover, recognizing the musical instrumental sounds, predicting the voice type, identifying the key chords(major/minor), and recommending music according to the classified genre.

1.2 Problem Domain

In the world, Music has a huge impact. From younger generations to the most senior generations music is an important part of life. Most people listen to music to entertain themselves to get into a certain mood, 90% of the world's population listens to music, according to (**Nielsen Music 360**). Out of the listening music percentage, a low percentage knows about music. (13% of people used to play an instrument and know music).

Music Genres

Music genres are the types of music that have specific distinguishable characteristics. They originated from different backgrounds and parts of the world. Basically pop, hip-hop, rock, jazz, metal, blues & disco are some of them. Modern music is originally based on these types of music. But it is hard to categorize modern music under genres though it has past features.

Musical Instrumental Sounds

Vibration, resonance, and amplification work together to create the sound that musical instruments make. The shape, size, material composition, and playing technique are only a few of the physical characteristics that each instrument possesses and which affect the sound it creates. It can be seen the ability of recognizing musical instrumental sounds (except guitar, violin, piano and drums) is low within the majority of music listeners.

Key and Chords

In theory of music the key is the group of pitches (frequency), or scale that forms the basis of musical composition of a piece of music, while a chord is a combination of 3 or more notes (pitches) which sounds simultaneously harmonically. There are 2 basic types of chords called major and minor and each having 12 key signatures. Categorizing chords as major and minor manually is harder for those who having lack of music knowledge.

1.3 Problem Definition

There is no proper system in the world that includes all the features (Music genre classification, audio classification - predicting voice color and instruments used, key/chord prediction, and music recommendation). As mentioned in the problem domain, people listening to music is increasing daily. But people knowing music or understanding music don't grow as music listening or streaming. So, the main problem is to reduce this gap. To reduce this gap the system will collect music, that will be classified into different genres. Instrumental sounds will be identified, voice range will be predicted, key/chord will be generated, and music will be recommended. To achieve this an AI-based application is required. Since it includes machine learning techniques and user inputs of music, the accuracy of this system will be increased, and it will minimize this gap.

1.3.1 Problem Statement

Ordinary music listeners do not have the ability or the training to identify most of the hidden details in a music track such as the genre, the musical instruments used, the voice type, and more which severely limits their ability to find similar music tracks they might enjoy.

1.4 Research Motivation

What is music? Music is a song made by different instruments. Music is a song made by voices with unique tone colors. Moreover, music is a set of sounds that harmonize with each other. Do people know these exactly? For this project, our primary motivation is to educate the basic things of music for normal people with our music knowledge. Developing this app will be a great way to inject good music taste among people around the world.

1.5 Existing Work

Table 1: Existing Work

	Limitations	Technology/ Algorithm	Advantages
Existing work on Music Genre Classification Systems			
(Shah et al, 2022)	Limitations in computational power and datasets containing a wider range of genres compelled the research to focus on simpler models instead of complex models for	CNN	A custom CNN architecture is used for the music recommendati

	music genre classification. Hence resulted in a relatively less accurate music genre classification system.		on systems which was very efficient in music feature extraction.
(Falola et al, 2022)	An outdated and relatively small dataset is used to train the ML model which could classify only 10 genres. This can cause limitations when predicting genres of modern music which has many new genres that are	Support Vector Machine (SVM)/CALM classifier	It was discovered that using 2 or more classifiers during the genre classification process delivered significantly better results.
Existing work on Music Recommendation Systems			

(Mesghali et al, 2022)	The music recommendation system utilizes past your behavior to recommend music that the user may be interested in but the recommendation model because inaccurate when the user is new and has no past behavior records.	Bi-LSTM	The Bi-LSTM model used in the research required significantly lesser training time compared to the traditional CNN model.
(Lee et al, 2015)	The efficiency of the music recommendation system is measured based on user behaviors of just 5 users which is the very small sample size and the accuracy of the recommendation system measured using the data generated by just 5 users is not very reliable.	MusicRecom (custom)	Research has a novel approach to music recommendation which utilizes both usage history and automatic genre classification to suggest highly personalized music to the user.
Existing work on Music Instrumental Sound Identification Systems			

(Yijie et al 2022)	The model requires further improvement in terms of detecting the instruments when multiple instruments are used in the same audio clip.	XGBoost	The accuracy of the algorithm in this paper on the public dataset medley-solos-DB reached 97. 65%.
(Zhang et al 2021)	The study focuses exclusively on traditional Chinese instruments and the models are not trained to detect sounds from modern musical instruments.	CNN/RNN/DNN	The model used in the research achieved a greater level of accuracy (82 – 99%)
Existing work on Voice Pitch Detection System			
(Rodrigo et al 2017)	The accuracy of the trained model declined when 4 or more voices were detected in the audio	PLCA-based acoustic model and an HMM-based music language model	A very efficient multi-pitch detection system with significant accuracy.

1.6 Research Gap

Due to the Advancement of technology, every past domain has been modernized into different platforms. As it is the same for the music industry. The advancement of technology has increased the crowd's engagement with music. Because of this people tend to use smartphones to listen to music. In recent years the music domain is a well-researched topic. Mainly music genre classification and audio classification are mainly pitched areas. As some researchers have done their relevant work on chord prediction and instrument predictions.

Every researcher has implemented models in predicting the above-mentioned features. The accuracy of the models is different, but the approaches used to classify the above features are the same (Using the same AI Algorithms). No researcher has not implemented a UI app (mobile or web).

So, during this research, the main pitched area was, researchers have just classified music into a genre and recommended music and vice versa. But the research gap proposed a new system that includes every mentioned feature in one user-friendly mobile application (The proposed system is based on mobile because music is listened to mostly using smartphones). The newly added feature which does not exist in any research in the proposed system is to predict the tone color of the user's voice/recorded clip.

1.7 Contribution to the Body of Knowledge

1.7.1 Technological contribution

In this system, implementation is done by processing the image spectrograms of the received music files. This means the images will be processed by AI algorithms of the system. The most feasible Algorithm is CNN (Convolutional Neural Network). This is to classify music genres, voice tone color, and prediction of the key and the instrument.

A separate daily updated database will be used to list down the trending music per day to recommend the music according to the generated genre. To get the most accuracy from our system the training models need to add or shuffle for the existing feature type.

1.7.2 Domain contribution

The project aims to contribute to the music domain in various ways by focusing on 20th and 21st-century music genres. The knowledge of music transformations will lead to an increase in the quality of music and its listeners. The app provides the feature of identifying key chords for music enthusiasts and classifies the tone color of audio into voice range types (soprano, alto, tenor, bass) and musical instrument-wise, a unique feature not found in other existing apps.

1.8 Research Challenge

During this project, we had to face some challenges. Our domain is a collection of main domains (Music Genre classification + Music recommendation + Chord/Key prediction + voice tone Prediction) so when searching for datasets we needed to look at datasets separately for all the domains and had to choose the most appropriate one.
When selecting a model, Our team had to face some difficulties. It is because we didn't have an idea about any model type. Our team took some time to find and get an agreement for our model.

1.9 Research Questions

R1. How to classify audio clips according to their features?

R2. How to recommend good music tracks according to the features of classified music?

R3. How to increase the accuracy and the quality of our product to give the best user experience?

R4. Which models can be used for the classification process?

1.10 Research Aim

The research aims to classify music information and give any music listener the much-needed basic knowledge about music.

1.11 Research Objectives

Table 2: Research Objectives

Research Objectives	Explanation	Learning Outcome
Problem Identification	Identify an efficient and accurate machine learning algorithm to detect and classify audio information.	LO1
Literature Review	RO1 – Identify a suitable programming language and machine learning algorithm to detect audio information. RO2 – Identify datasets to train the Multimodal Music Information Classification System RO3 – To design the system to detect music tracks from a plain audio. RO4 – To design the system to further detect the tone and genre and temper of the music track. RO5 – To design the system to suggest music recommendations to the user based on the data gathered in the previous steps. RO6 - Implement the Multimodal Audio Classification System using Mobile/Web Application using an intuitive UI.	LO1
Data Gathering	Utilize the datasets of music information with the help of online data libraries.	LO2, LO3

and Analysis		
Research Design	Identify the best dataset to train the Multimodal Music Information Classification System from amongst existing publicly available datasets.	L04
Implementation	Implement the Multimodal Music Information Classification System using a mobile/web application.	L04
Testing and Evaluation	Identify the accuracy of the Multimodal Music Information Classification System using user feedback.	L04

1.12 Project Scope

1.12.1 In- Scope

Table 3: In-scope

No	Description
1	Identifying selected music genres.
2	Predicting the tone color of the voice. (SATB)
3	Newly trending music will be recommended according to the time.
4	Identifying the Key Chord of the music.
5	Listing the musical instruments used to create the music.

1.12.2 Out-scope

Table 4: Out-scope

No	Description
1	Classifying Sri Lankan Music into genres.
2	Every audio file cannot be tested.
3	Genres are limited only the most listened genres are tested.
4	Chords will not be generated.

1.12.3 Feature Prototype

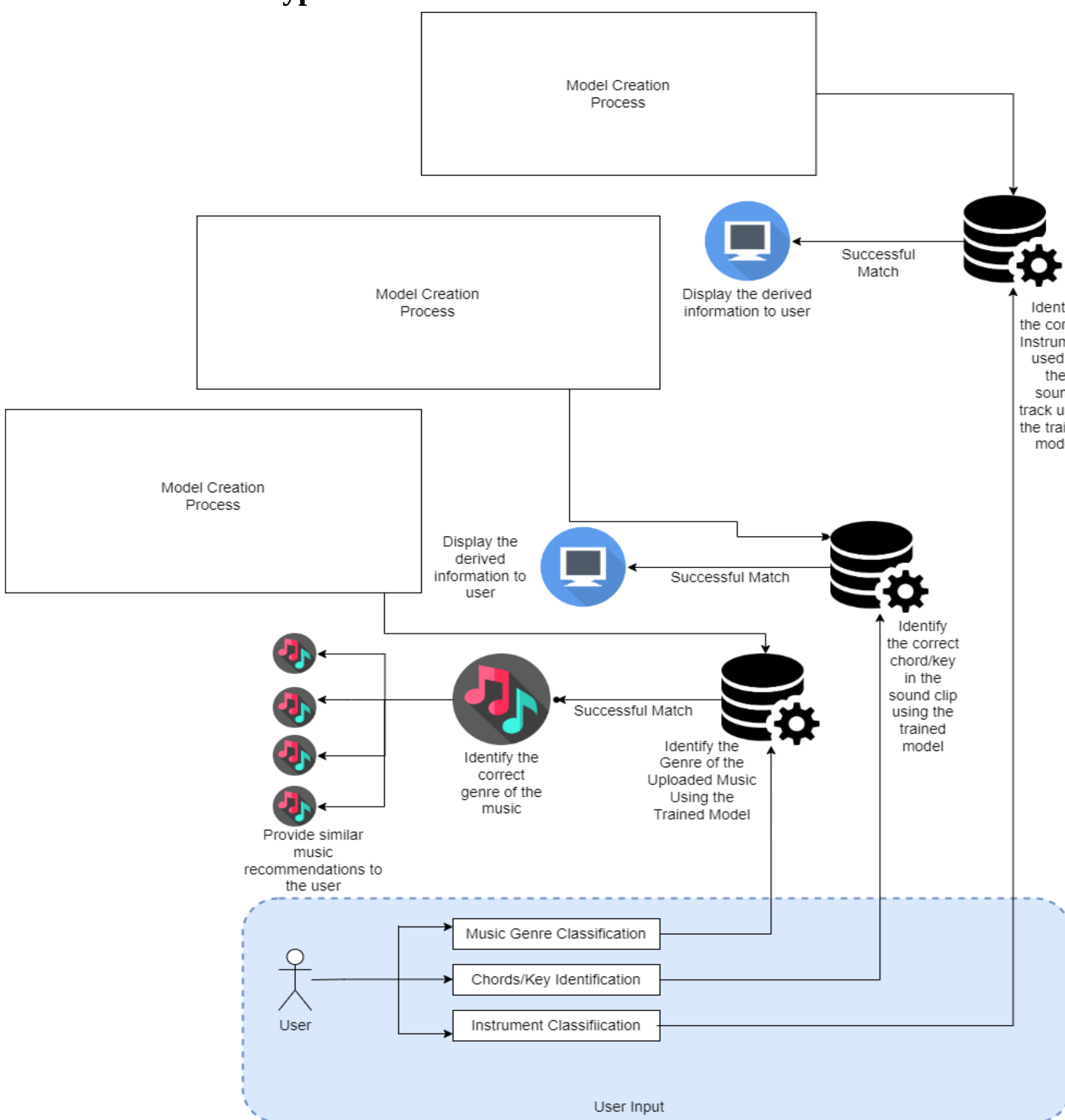


Figure 1: Prototype Diagram

1.12 Resource Requirements

1.13.1 Hardware Requirements

- Central Processing Unit (CPU) — Intel Core i5 6th Generation processor or higher. An AMD (Advanced Micro Devices) equivalent processor will also be optimal. To train and run the model.
- RAM — 8 GB minimum (7.78 GB usable), 16 GB or higher is recommended. To image processing images, create the spectrograms and store these images.
- Graphics Processing Unit (GPU) — NVIDIA GeForce GTX 960 or higher.

1.13.2 Software Requirements

- Python - The main programming language to build is python and it also supports many libraries.
  - TensorFlow – Time series analysis, speech, and image recognition, and training of the model.
  - Numpy – Working with arrays.
  - Scipy - Multidimensional image processing.
  - Pandas - Easy to deal with missing data.
  - Matplotlib - Low memory consumption and better runtime behavior.
  - Magenta, Librosa, and pyAudioAnalysis – working with audio data.
- MS Word – For creating documentation.
- Google Drive – To store huge-size data sets and other valuable documents.
- Windows operating system – To handle huge computational functionalities.
- Git - Version controller and easy to work with a team.
- Flutter - To create UI in the mobile-based application.
- Anaconda Navigator - Combining all functionalities into one place and using

python IDE as Jupiter notebook.

- Draw.io/star UML – To design wireframes for the proposed system.
- Jira – Project management tool.

1.13.3 Skills Requirements

Technical skills – Programming, knowledge of Analytical tools, processing large data sets and machine learning and deep learning, and statistical analysis.

Non-technical skills – Strong business acumen, and communication skills.

1.13.4 Data Requirements

This project includes 4 mini-projects with a machine-learning approach. Therefore 4 datasets are required. GTZAN was the most used dataset for music genre classification and will be used here. It includes 1000 audio files with a duration of 30 seconds each. The Cantorina dataset will use for vocal range identification. IRMAS: a dataset for instrument recognition in musical audio signals will be another dataset. Music Chords (Major/Minor) dataset with chord sounds of 859 audio clips will be used to identify key chords of the music.

1.14 Chapter Summary

The music industry has undergone many changes in the past decade with the introduction of music streaming platforms. Despite all the advances in music streaming technologies, the public seem to have little to no knowledge of different genres, musical instruments, vocal tones and pitch used in a music track hence limiting their ability to pinpoint what they like or dislike in a given music track. Thus, there is a tremendous interest among the music listeners for a centralized, intuitive audio and music classification system. Recent advances in machine learning algorithms have made it feasible to create such a system that is affordable, accessible.

2. CHAPTER 02: LITERATURE REVIEW

2.1 Chapter Overview

This chapter provides an overview of the current state of the art in the field of audio classification. This chapter will explore various research studies, academic articles, and other relevant sources of information related to audio classification systems. The review will cover the latest advancements in audio classification techniques, including machine learning algorithms, and their application in multi-modal classification systems.

2.2 Problem Domain

In the modern era, people are finding a way to relax by listening to music on their smart devices.

Commonly their approach would be downloading music tracks and categorizing them according to their choice into playlists or refusing that if they don't like it. Here as the project mainly aims to help people to classify the music. Moreover, classification of audio track (instrumental / voice/ voice + instrumental), classify the music according to the genre, and recommend more music for the user's preference. Hence, a user could input the list of audio files and get the output including the features of the tracks and genre of music.

2.3 Existing Work

2.3.1 Audio Classification

In the past few years, several approaches have been taken for audio classification using machine learning and deep learning. An audio recognition system for Chinese traditional instruments was made, with Mel spectrum characteristics as input. (Rongfeng Li et al., 2022) has trained an 8-layer convolutional neural network.

Instead of reducing model size using complex methods, (Xubo Liu et al., 2022) has improved a model to eliminate the temporal redundancy in the input audio features (Mel spectrogram) and proposing a family of simple pooling front ends (SimPFs) to reduce redundant information within Mel-spectrogram. Another approach was by (Jianyuan Sun et al., 2022) using deep neural decision forest (DNDF) to classify an audio clip based on the characteristics of the recording environment (ASC). DNDF combines a fixed number of convolutional layers and a decision forest as the final classifier.

There's an approach considering the number of non-target events (Wu Dan, 2019) and problems such as detection strategy, detection time, the decision tree models. (Liang Gao et al., 2022) did 2 multiple representations as inputs to train the networks jointly with a knowledge distillation method.

2.3.2 Music Genre Classification

There have been Several research done to categorize music. Up until now, many techniques were investigated for music genre classification. A multi-frame approach (Tejas Dalvi et al.,2022) with an average stage to analyze in detail almost the full song. It is used at training time to generate more samples and at test, time to achieve an overview of the whole song. The model to evaluate the performance of the multi-frame approach has been trained with the train partition of hand-made dataset and evaluated using the test partition.

A Convolutional Neural Network (CNN), a deep learning technique was created by (Peace Busola et al., 2022) with a total accuracy of 77.89%. The deep learning approach made by him could play a vital role in classification, audio features such as spectrogram features which are extracted from the signal of the music are one of the best features that gave excellent results.

Another research was conducted by (Hongjuan Zhang et al.,2022) where they tried identifying a novel classification framework incorporating the auditory image feature with traditional acoustic features and spectral features which proposed to improve the classification accuracy. Moreover,the logarithmic frequency spectrogram (Dhevan Lau al.,2020) rather than linear is adopted to extract the spectral feature to capture the information about the low-frequency part adequately.Other machine learning and deep learning models can still be worked upon for accurate music genre classification. One of the most common datasets used for music genre classification is the GTZAN Database (Hansi Yang et al., 2019), (Lata Gohil et al., 2022) and (Yeshwant Singh et al.,2022)

.

2.4 Technology/Approach/Algorithm Review

There are various technologies, approaches, and algorithms used in existing audio classification systems. These systems typically use machine learning algorithms such as support vector machines (SVM), decision trees, and neural networks to classify audio files. Features extraction is a crucial step in audio classification, and different methods like mel-frequency cepstral coefficients (MFCC), linear predictive coding (LPC), and wavelet transforms are used to extract relevant features from audio signals. Recent advances in deep learning, such as convolutional neural networks (CNN) and recurrent neural networks (RNN), have also shown significant promise in improving the accuracy of audio classification systems. In addition, multi-modal approaches, which integrate data from multiple sources such as audio and video, have been shown to improve the performance of audio classification systems. Overall, a combination of these technologies, approaches, and algorithms can be used to create an effective audio classification system.

2.5 Chapter Summary

There are some implementation-related issues with earlier studies focused on music genre categorization, audio classification, and music recommendation. When attempting to categorize music genre, the models employed to extract features from audio had several limitations (could not interpret every nuance of scatter spectrograms). There were issues with how to categorize audio features from image spectrograms when training the models. However, the number of participants utilized to assess the accuracy of the models used in these studies was tiny, making it difficult to trust the accuracy of the models used to read, categorize, and recommend music.

Also, the lengthy machine learning models used in these studies for categorizing and recommending music can lead to a poor user experience. Multimodal Music Information Classification System is the focus of our project. The audio files that will be used as data are wav files with a maximum duration of 30 seconds and midi files. The audio is categorized and filtered by search algorithms. If the audio files are solely music tracks, then machine learning techniques will be used to classify them into the appropriate genres after audio detection, and later music will be suggested as per feature. The research gap mostly focuses on combining the above three components into a single mobile application.

3. CHAPTER 03: METHODOLOGY

3.1 Chapter Overview

This chapter outlines the methods used to conduct the research. It provides a detailed description of different methodologies followed to conduct the research, development, design, evaluation, and project management. Furthermore a expected timeline for the project implementation can be found.

3.2 Research Methodology

Table 5: Research Methodology

Research Philosophy	Positivism is chosen as the research philosophy for this project as it involves collection of accurate data, identifying objective patterns in the data using statistics, and classifying them accordingly.
Research Approach	Quantitative approach will be used as it involves gathering larger samples of data and usage of statistical methods to analyze the data.

Research Strategy	Experiment Research strategy will be followed in this project as it involves manipulating one or more variables to see how they affect the outcome.
Research Choice	A Multi method research will be chosen as it can help overcome limitations and provide a more comprehensive understanding of the research question.
Time zone	The research uses a pre-defined data set that is collected over a large time.

3.3 Development Methodology

What is the life cycle model and why?

As agile project life cycle model allows to make rapid changes based on outcome and feedback of the client/supervisor. The Agile model allows greater flexibility and improves collaboration among members and greater quality of the final product while enabling the team stick to implementation timeline.

Design methodology => SSADM or OOAD or Anything else?

Object-Oriented Analysis and Design (OOAD) is chosen over the Structured Systems Analysis and Design Method (SSADM) due to the complex nature of the project and the need for flexibility during the design and implementation process. The OOAD approach enables a more iterative and incremental approach, which allows the development team to make changes and adjustments to the system as it is being built. The OOAD approach also allows individual members to work on various components of the project and integrate them to the main system.

Evaluation methodology => Evaluation metrics and/or benchmarking.

The F1 score is a relevant evaluation metric for this project because the system is anticipated to categorize music tracks into various genres, recognize keys and chords, and identify the musical instruments utilized in each audio track. The F1 score is a single number that represents the system's performance in terms of recall and precision. It can be used to compare the effectiveness of various models.

3.4 Project Management Methodology

3.4.1 Schedule using Gantt Chart after doing WBS

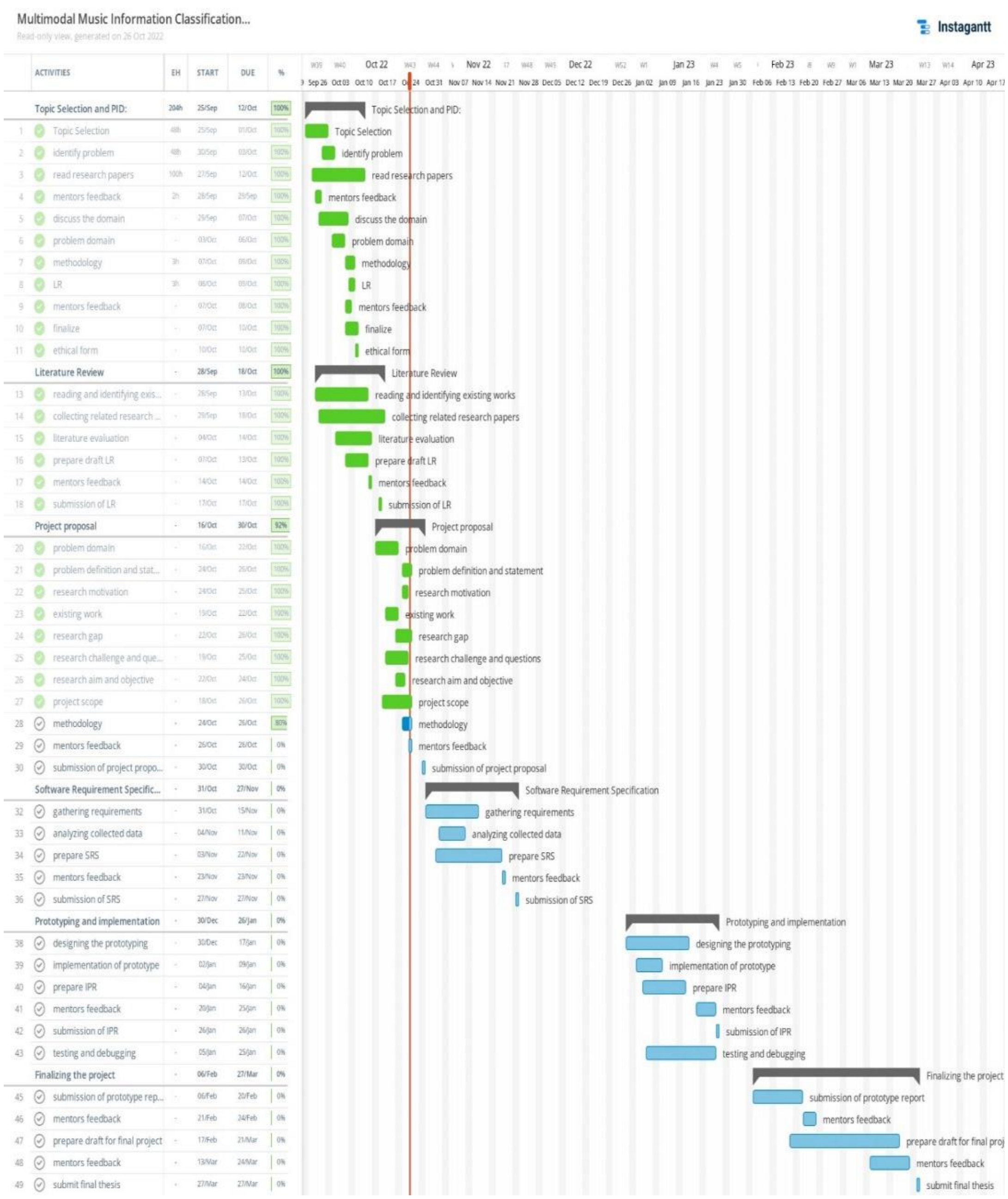


Figure 2: Gantt Chart

3.4.2 Deliverables

Table 6: Deliverables

Deliverable / Milestone	Due Date
Datasets and Components of the project	2nd week
Project Proposal	5th week
SRS	9th week
Learning the tech stack	9th week
Building the components of the project	17th week
Testing	19th week
Integration Process	19th week
Develop CI/CD Pipelines and Integrate to each component	21st week
Final Testing Phase of the applications	23rd week
Evaluation Phase	24th week
Post-Mortem and Research Paper	24th week

3.5 Chapter Summary

Development, design, evaluation, and project management methodologies discussed in this chapter ensures the quality, accuracy, and reliability of the multi modal music classification system. The evaluation metrics selected to measure the accuracy of the system help to determine the success of the implementation.

4. CHAPTER 04: SOFTWARE REQUIREMENTS SPECIFICATION

4.1 Chapter Overview

The major goal of this document is to identify the important stakeholders. to collect the needs and assess the requirements to determine the information that should be given priority. A detailed picture of the project's internal and external settings as well as an overview of its stakeholders is first drawn. To evaluate the various project techniques, a requirement elicitation was performed. In the last section, functional and non-functional requirements are described along with a use case diagram and its use case description.

4.2 Rich Picture

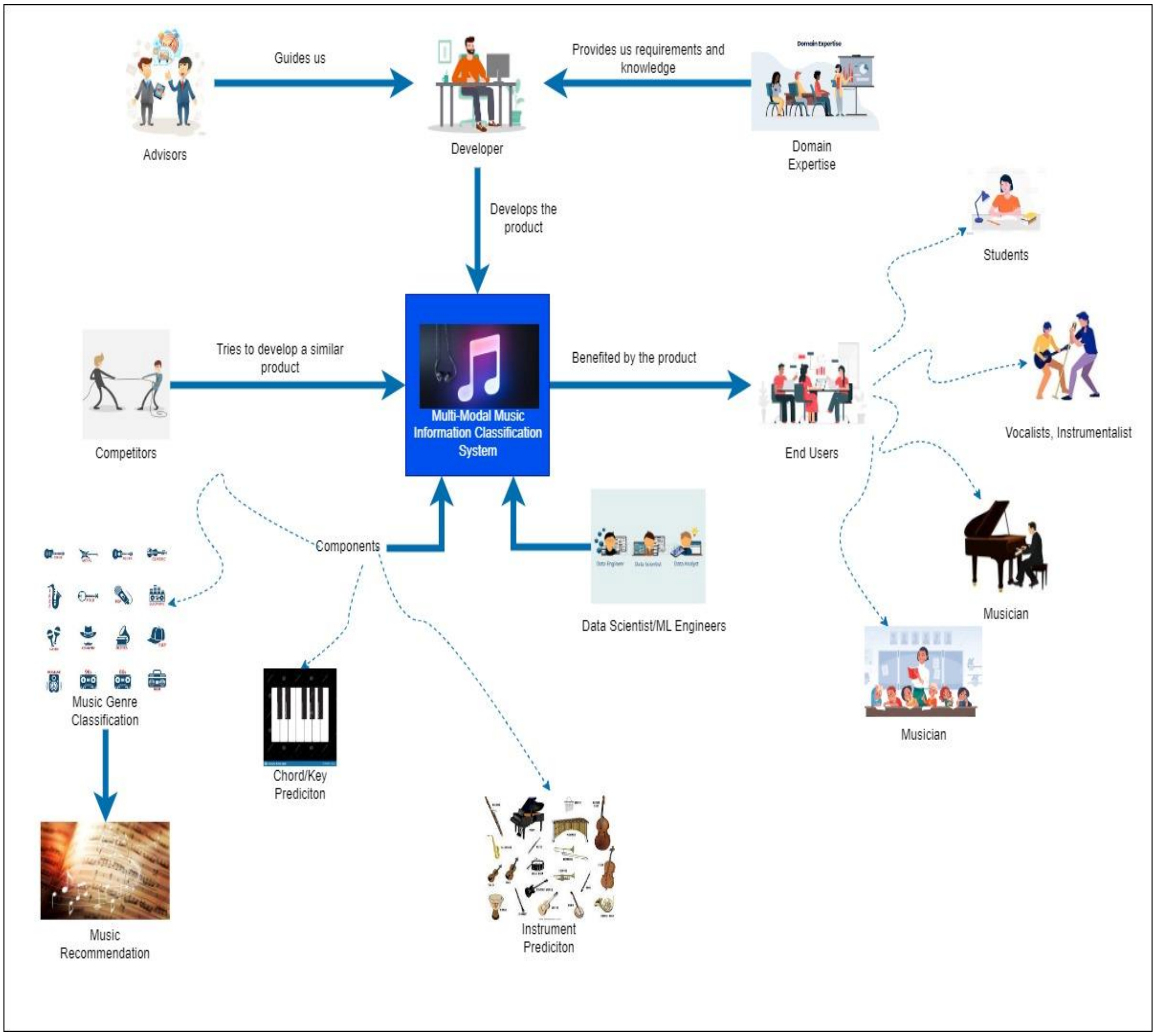


Figure 3: Rich Picture

The above rich picture illustrates the different stakeholders engaged with the Multimodal Music Information Classification (MMIC) System for different purposes. Advisors, developers, domain experts, competitors, end users, and data scientists/machine learning engineers are the identified stakeholders.

4.3 Stakeholder Analysis

In the below, the associated stakeholders are shown by the onion model diagram. The role and viewpoint of each stakeholder are identified.

4.3.1 Onion Model

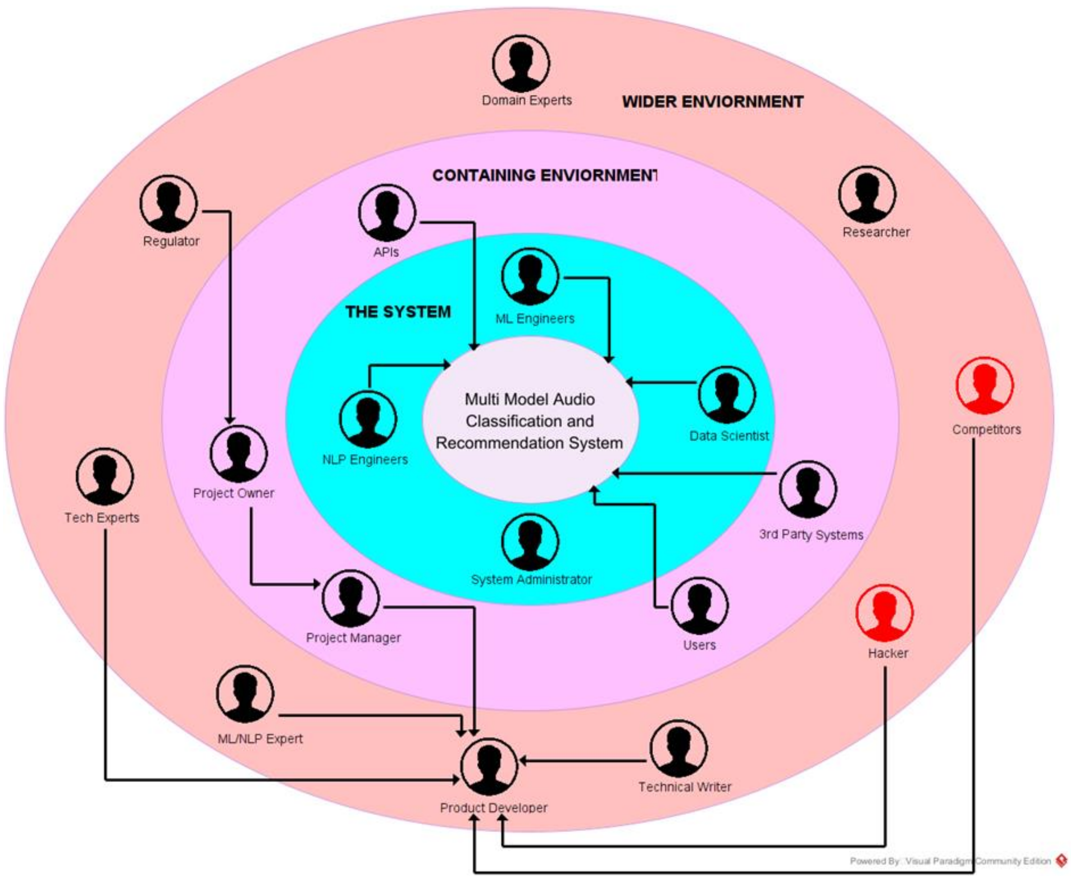


Figure 4: Onion Modal

4.3.2 Stakeholders Viewpoints

Table 7: Stakeholder's Viewpoints

Stakeholder	Role	Benefits
Data scientists, ML Engineers	Operational Maintenance	Design and development of the MMIC System process and models.
System Admin	Operational Administration	Deploying and configuring the application for various environments.

Product Owner	Functional Beneficiary	Owner of the MMIC system.
Product Manager	Managerial Support	Managing the application and processes to ensure that the project works properly.
Users, 3rd Party Systems, APIs	Functional Beneficiary	Using the developed MMIC application through various channels (Android) or integrating it with other systems.
Regulator	Quality Regulator	Monitor the application to make sure that data usage and processing fall within the established privacy policies.
Tech Experts	Expert	Determines if a collection of specifications is supported by the platform.
Domain Experts	Expert	Provides a domain view of the project's technology and methodologies.
Product Developer	Developer, Operational Maintenance	Implements and maintains the system.
Technical Writer	Operational Support	Provides support on the creation of the documentation for the system.
Researcher	Educational Beneficiary	Analyses the current systems and approaches to increase the efficiency of the current process and techniques.
Competitor	Negative Stakeholder	Creates an application that has similar features to the proposed system.
Hacker	Negative Stakeholder	Poses a threat to the proposed system due to the possibility of unauthorized access to the application and its data.

4.4 Selection of Requirement Elicitation Techniques

The various methods of gathering requirements are referred to as requirement elicitation. Requirement Elicitation Methods are employed with the assistance of clients, users, and other stakeholders to ascertain the system's requirements. The best strategies taken into consideration are described in the sections that follow, along with some of their benefits and drawbacks.

4.4.1 Observing Existing Systems and Literature Review

Reviewing current systems that have similar features and approaches may be the initial step for requirement elicitation. The current work in the domain is classified and analyzed to identify features that could be improved.

Table 8: Pros and Cons of existing systems and literature review

Advantages	**Disadvantages**
·The main components for MMAC system, approaches for implementation can be identified. ·Useful in identifying the feature gaps which can be further improved.	·Even though identifying the features of the available commercial products are not complex, reviewing the existing research and developed systems are complex, since the objective of each research is different.

4.4.2 Surveys & Questionnaires

The MMAC system can used by potentially anyone who is interested in music, or regular music listener who has access to internet, a computer or even a smartphone. Since MMAC system has a very broad target audience, using surveys and questionnaires is ideal as a wide audience can be covered and large amount of reliable data can be gathered.

Table 9: Pros and cons of survey and questionaries

Advantages	Disadvantages
• Ability to cover a very large, diversified audience. • Less time consuming and can be completed with ease by anyone. • Easy to classify and analyze survey data.	• Multiple choice questions can be tricky as it is quite hard to know exactly what a person feels about a question. • The question can be misunderstood, or the answers may not be honest.

4.4.3 Interviews

Interviews can provide an insight into current gaps and opportunities for improvement in the domain for the MMAC system. Experts in the music industry, regular music listeners and students who study music could be interviewed for this purpose.

Table 10: Pros and cons of interviews

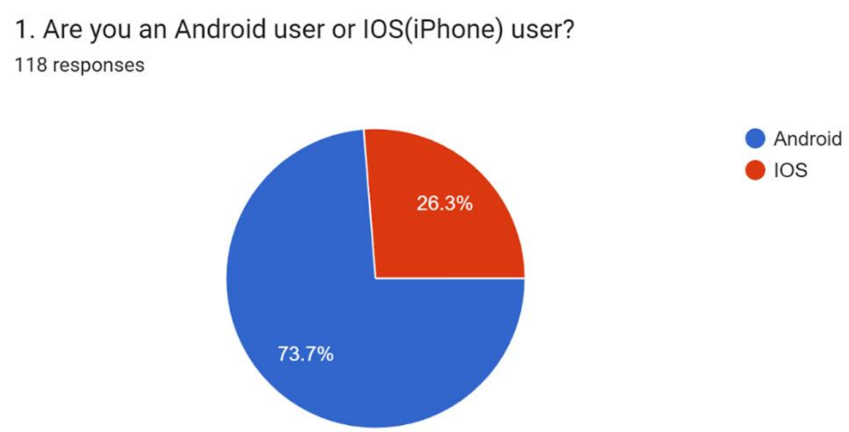
Advantages	Disadvantages
• Direct face to face interactions can provide unique point of view of the problem. • The face-to-face interview reduce the chances of misinterpretations compared to surveys and questionnaires.	• Covering a wide diversified audience is challenging and time consuming. • · Everyone may not have to the time to allocate for the interviews and may not have enough knowledge to understand the tech used in the MMAC system.

4.5 Discussion of Results

4.5.1 Survey Findings

Table 11: Survey Findings

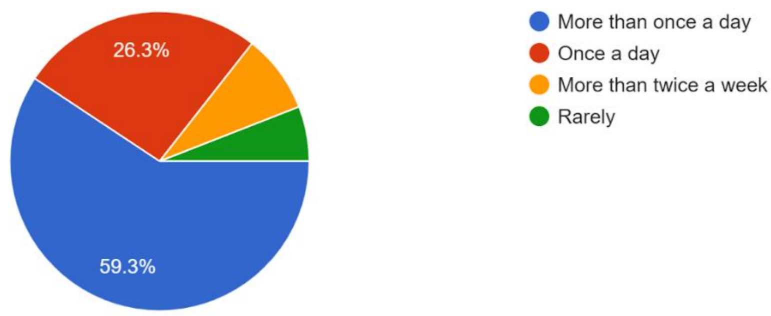
Question	Are you an Android user or IOS (iPhone) user?
Aim of the Question	To see the number of users who use IOs and Android operating systems.
Observations	



The observation is analyzed from 73.7% of Android users and 26.3% of IOs users.

Conclusions	Most of the participants are Android users. So, the product is an Android OS-based mobile application.
Question	How often do you listen to music?
Aim of the Question	To see how users listen to music.
Observation	

2. How often do you listen to music?
118 responses

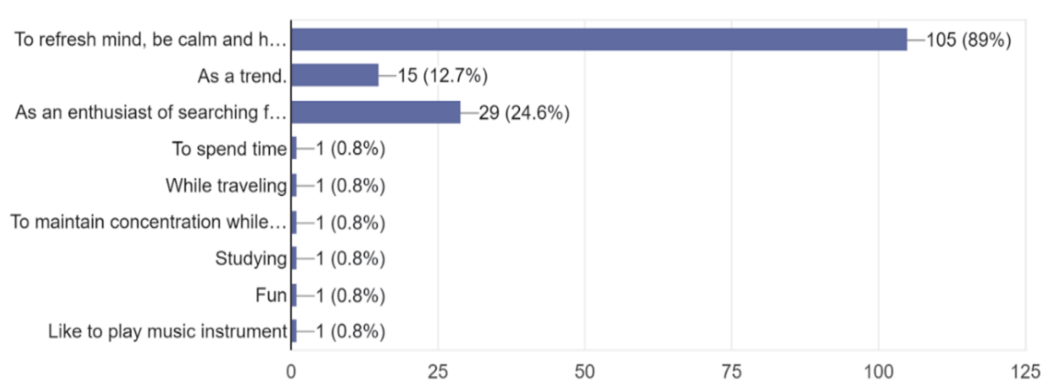


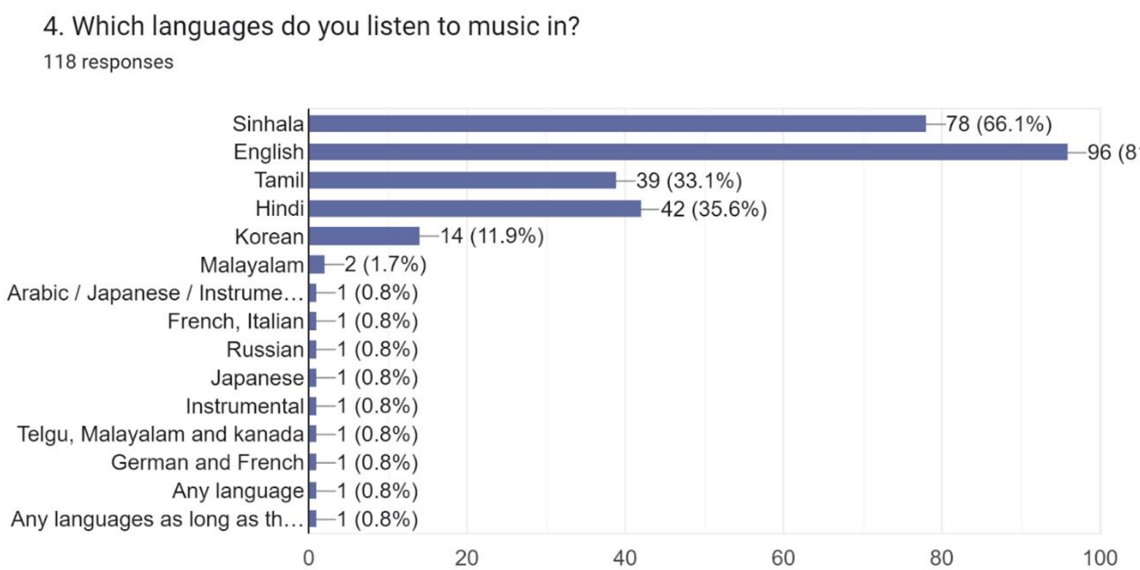
It is observed that 59.3% of participants listen to music more than once a day, 59.3 % once a day and 15.4% listen to music twice a week and rarely.

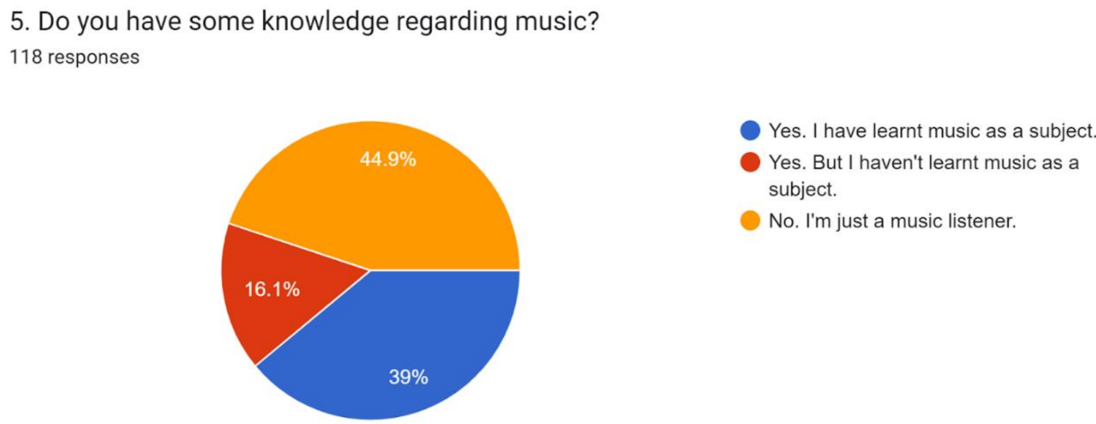
Conclusion	It is observed that most of the participants listen to music frequently and a least number of participants neglect music.
Question	Why do you listen to music?
Aim of the Question	To analyze why the users, listen to music.

Observation

3. Why do you listen to music? (If you have a special reason mention it on Other)
118 responses

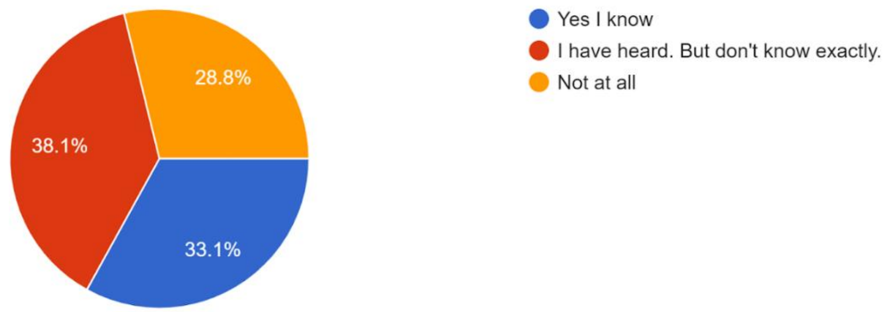


It is observed that 89% of the participants listen to music to freshen and calm themselves to release their stress of work. 24.6% listen to music for their passion for learning music and 12.7% listen to music as a trend in this society.	
Conclusion	Today, due to the high workload and survival of life people are stressed. So, the only remedy to get rid of it is to listen to music and most people have a lesser knowledge of music.
Question	Which language do you listen to music in?
Aim of the Question	To identify the taste in music they listen to.
Observation  The top 3 languages that users listen to are English, Sinhala and Hindi. Out of them English has the highest percentage of 81.4%.	
Conclusion	English is the most listened to type of music. So that the product is an English-only classifying system.

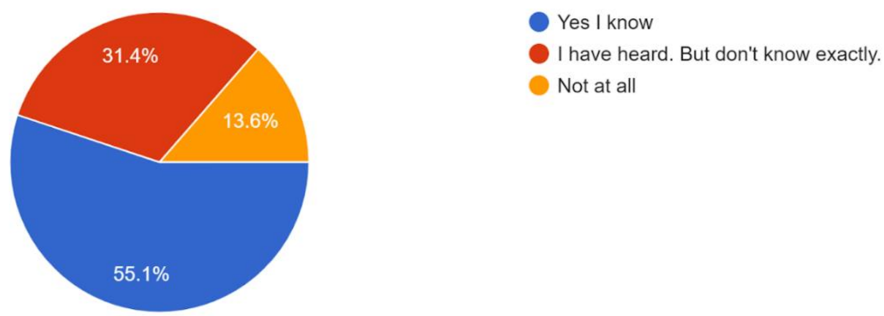
Question	Do you have some knowledge regarding music?
Aim of the Question	To see the participants' knowledge of music.
Observation  44.9% of participants are just music listeners, 39% have learnt music as a subject and 16.1% haven't learnt music as a subject.	
Conclusion	This product will be implemented to reduce the gap of music listeners to convert them to music knowledgeable listeners.
Question	Do you know about voice ranges in music?Do you know about different ranges of music? Do you know about key signatures?
Aim of the Question	In this question the main objective is to identify the different knowledge about music.

Observation

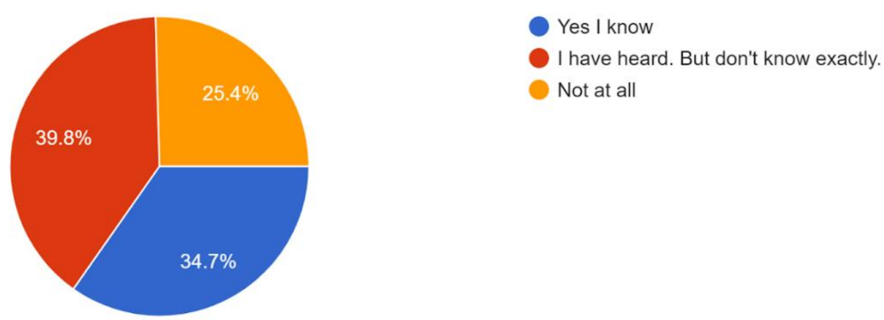
6. Do you know about voice ranges in music? (Soprano, Alto, Tenor, Bass)
118 responses



7. Do you know about different music genres? (Disco, Reggae, Blues, Jazz, etc.)
118 responses



8. Do you know about key signatures and chords in music?
118 responses



Out of the three questions most of the survey participants have heard about music genres (31%), voice ranges (38%) and key/chord (40%).	
Conclusion	The participants have a minimum knowledge about the above questioned categories.

4.5.2 Interviews

Interviews were conducted with domain experts to discuss the scope of this project. Mainly we met and interviewed parties who engage in the music industry. We mainly discussed the gap in knowledge of music and how this product can be utilized to decrease this gap. Mainly discussed topics are,

Table 12: Discussion of Interview Results

Question 1	Present society's knowledge and passion about music?
Conclusion	People in this society have less of an impression and a passion to learn music compared to the last decades.
Question 2	What reason for this drop in passion?
Conclusion	Due to the increase in living costs people are struggling to work on their aesthetic taste and the increase in technology. The younger generation tends to play video games without learning an art subject.
Question 3	Will the proposed system help this gap to be reduced?
Conclusion	As this system is a technological-based simple mobile application it is a well-adaptable system. So, there is a positive aspect that this product will reduce this gap.

4.6 Summary of Findings

Table 13: Summary of findings

Findings	Literature Review	Questionnaire	Existing Systems	Interviews
A system with all mentioned components is not there.	X	X		X
Most people have preferred a mobile application with Android as its OS.		X		
Using ML and DL techniques is necessary for implementing the components	X			
Knowledge of music is considerably less		X		X
A music classification system will help to increase the knowledge of music				X
No chord key prediction systems which use audio files to predict.			X	
Apps that classify music into genres do not use the	X		X	

input to recommend music.				
Music Recommendation Systems recommend music according to user preference only.	**X**		**X**	

4.7 Context Diagram

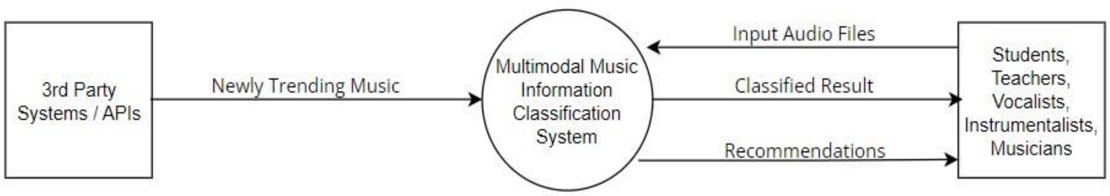


Figure 5: Context Diagram

4.8 Use Case Diagram

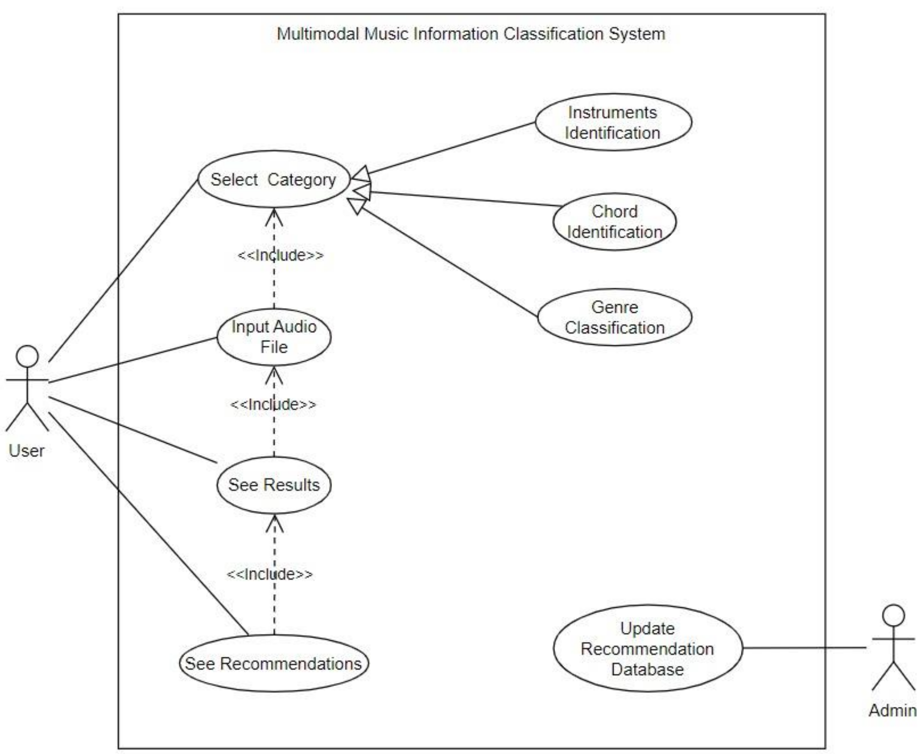


Figure 6: Use case Diagram

4.8.1 Use Case Description

Table 14: Use case description

<table>
<tr><td>Use Case ID</td><td>UC001</td></tr>
<tr><td>Use Case Name</td><td>Select Category</td></tr>
<tr><td>Actor</td><td>User</td></tr>
<tr><td>Purpose</td><td>Select a category from 3 categories in the main interface.</td></tr>
<tr><td>Overview</td><td>The user needs to select a category.
The user selects a category from the main interface.
The user directs to the selected category interface.</td></tr>
<tr><td>Pre-Conditions</td><td>The user should open the app.</td></tr>
<tr><td>Post Conditions</td><td>The user should be in the selected category interface.</td></tr>
<tr><td>The Typical Course of Events.</td><td><table><tr><td>Actor action</td><td>System response</td></tr><tr><td>The user selects a category from the main interface.</td><td>The system directs the user to the selected category interface.</td></tr></table></td></tr>
<tr><td>Alternative Courses</td><td>-</td></tr>
</table>

<table>
<tr><td>Use Case ID</td><td>UC002</td></tr>
<tr><td>Use Case Name</td><td>Input Audio File</td></tr>
<tr><td>Actor</td><td>User</td></tr>
<tr><td>Purpose</td><td>Input the audio clip to get the relevant output.</td></tr>
<tr><td>Overview</td><td>The user needs to classify music genre/voice type/instruments or identify key chords on a music track.
The user selects the music track from the device storage.
The user inputs the audio file into the system.</td></tr>
<tr><td>Pre-Conditions</td><td>The user should be in the selected category interface.</td></tr>
<tr><td>Post Conditions</td><td>The user should successfully input the audio file into the system.</td></tr>
<tr><td></td><td><table><tr><td>Actor action</td><td>System response</td></tr></table></td></tr>
</table>

The Typical Course of Events.	The user gives permits to access device files. The user selects a music track from the device storage.	Access device files. Copy the selected audio clip into the system.
Alternative Courses	Unable to copy larger files or incorrect file formats into the system.	

Use Case ID	UC003	
Use Case Name	View Results	
Actor	User	
Purpose	View the results.	
Overview	The user needs to view the result. The user asks to show the result. The user views the result.	
Pre-Conditions	The user should have input the audio file successfully.	
Post Conditions	The user should see the result.	
The Typical Course of Events.	Actor action	System response
	The user clicks on ‘view results	The system runs the audio track and shows the result.
Alternative Courses	Unidentified audio clip.	

4.10 Functional Requirements

The functional requirements of the system are listed in the table below, along with their priority level.

Table 15: Functional Requirements

Priority Level	Description
Critical	Main features and functionalities must be mandatorily included in the system.

Important	A proposed feature or functionality that will further add value to the system but is not mandatory.
Desirable	An out-of-scope requirement

	Requirement and Description	**Priority**
FR01	Accepting the audio track	Critical
	The system should be able to accept the audio track the user wants to analyze.	

FR02	Determine if the audio track is a song with music, music only, or a voice track without any music	Critical
	The system should be able to pre-classify the user input before selecting the ideal model for further analysis of the audio track	
FR03	Detect the genre of the audio track	Critical
	The trained ML model for detecting music genre is used to detect the genre of the music track	
FR04	Detect the tone of the voice type of the track input by the user	Critical
	The trained ML model for detecting voice type is used to detect the voice type used in the track.	
FR05	Detect the instruments used in the music track	Critical
	The trained ML model for detecting voice type is used to detect the music instruments used in the track.	
FR06	Identify key chords used in the music track	Critical

	The trained ML model for detecting key chords is used to detect the music instruments used in the track.	
FR07	Recommendation of similar music	Important
	Based on the genre detected by the model, the system should be able to recommend similar music to the user.	
FR08	Save the results of the analysis to further improve the models	Critical
	Despite not having any use for the end user, metadata generated by each operation should be saved to further improve the ML models.	
FR09	GUI and other Interface support	Desirable
	A user-friendly intuitive UI and API interfaces for the 3rd party apps to connect with the system.	
FR10	Continuous training optimization of the systems	Desirable
	Models in the system should learn to improve the accuracy of detecting different features in the audio tracks.	

4.11 Non- Functional Requirements

Performance constraints – Reliability, security, response time, performance

Interface constraints - Usability

Economic constraints – Marketing analysis, resources, development cost

Lifecycle requirements – Quality of the design, solution evaluation, portability.

Table 16: Non-functional requirements

	Requirement and Description	Specification	Priority
NFR01	The accuracy of the system in detecting various features in an audio track should be high	Accuracy	Important
NFR02	The system should be secured to avoid unauthorized access and data breaches.	Security	Important
NFR03	Model training should be efficient and not take a long time	Performance	Important
NFR04	The process should be done with minimum resource requirements to support as many devices as possible.	Performance	Non-important
NFR05	The system will have an intuitive GUI	Usability	Desirable

4.12 Chapter Summary

An onion model and a thorough system analysis were presented in the chapter's opening paragraphs. A questionnaire result and specified requirement elicitation methods were also provided. A context diagram and a use case diagram were also used for demonstration. Moreover, requirements that were both functional and non-functional were listed with priorities.

5. CHAPTER 05: SOCIAL, LEGAL, ETHICAL AND PROFESSIONAL ISSUES

5.1 Chapter Overview

This chapter gives a comprehensive assessment of all the social, legal, ethical, and professional challenges that might affect the final result. These external influences can be examined and handled using the SLEP analysis. As a result, SLEP analysis is utilized to pinpoint potential problems that could occur; methods taken to address these problems are then covered in more detail in this chapter.

5.2 SLEP issues and mitigation

5.2.1 Social Issues

- Diversity and Inclusion: The system is designed to accommodate a wide range of musical styles and cultural perspectives eliminating any bias towards certain musical styles.
- Bias: The models are trained to detect a diverse range of musical genres and instruments eliminating the potential bias towards certain genres or artists.
- Language: English is used as the primary language of the product due to its universality, but this does not hinder the ability to classify or analyze audio tracks that may belong to languages other than English.
- Cultural appropriation: The system is trained in a diverse dataset using automated classification techniques to overcome misclassification and to ensure that music is classified accurately and respectfully.

5.2.2 Legal Issues

- Copyright infringement: The system is designed with a complete database of music metadata that includes information about copyright ownership and licensing. The system is built in such a way that illegal download or streaming of music is prevented.
- Data protection: User data is not sold or distributed to 3rd parties. The system is transparent about how the user data is collected, stored, and used.
- Dataset: All the datasets used in the training of the models are available in the public domain and are obtained with relevant permissions.

5.2.3 Ethical Issues

- Responsible data use: The system is built with responsible data use practices, with a focus on data privacy, security, and user control.
- Anonymity: The information gathered during surveys and interviews are stored anonymously to protect the identity of the participants.
- Bias and Discrimination: The system uses a diverse range of data in training the models to prevent misclassification and bias towards certain music types.
- Datasets: Datasets used in the training of models are chosen after ensuring that the data is gathered by ethical means and does not infringe upon the copyright of any 3rd party

5.2.4 Professional Issues

- Performance and reliability: The system is designed by adhering to wide accepted design principles making updating and integration seamless provide a positive user experience.
- Scalability and maintenance: Given the huge target audience the system is designed with scalability and ease of maintenance in mind. A complete documentation and testing process is in place to ensure that new features and updates can be integrated seamlessly.
- Collaboration and communication: The system is designed and updated by the team using version control such as git and GitHub. In addition to that task flow is managed using Trello and Email and WhatsApp is used primarily for communication.

5.3 Chapter Summary

This chapter provides an overview of the possible social, legal, ethical, and professional issues and how they might affect the system and steps taken for mitigation of these issues that may arise.

References

Shah, M., Pujara, N., Mangaroliya, K., Gohil, L., Vyas, T. and Degadwala, S. (2022). Music Genre Classification using Deep Learning. 2022 6th International Conference on Computing Methodologies and Communication (ICCMC). doi:10.1109/iccmc53470.2022.9753953.

Falola, P.B., Alabi, E.O., Ogunajo, F.T. and Fasae, O.D. (2022). MUSIC GENRE CLASSIFICATION USING MACHINE AND DEEP LEARNING TECHNIQUES: A REVIEW.ResearchJet Journal of Analysis and Inventions, [online] 3(03), pp.35–50. doi:10.17605/OSF.IO/FZQXW.

Dalvi, T., Kamble, S. and Yadav, S. (2022). MUSIC GENRE CLASSIFICATION. International Journal of Research Publication and Reviews Journal homepage: www.ijrpr.com, [online] 3, pp.2029–2034. Available at: https://ijrpr.com/uploads/V3ISSUE4/IJRPR3596.pdf [Accessed 15 Oct. 2022].

Cai, X. and Zhang, H. (2022). Music genre classification based on auditory image, spectral and acoustic features. Multimedia Systems. doi:10.1007/s00530-021-00886-3.

1Mesghali, S. and Askari, J. (2022). Designing Music Recommendation System based on Music Genre by using Bi-LSTM. [online] IEEE Xplore. doi:10.1109/ICEE55646.2022.9827016

Lee, J., Shin, S., Jang, D., Jang, S.-J. and Yoon, K. (2015). Music recommendation system based on usage history and automatic genre classification. [online] IEEE Xplore. doi:10.1109/ICCE.2015.7066352

Liu, Y., Yin, Y., Zhu, Q. and Cui, W. (2022). Musical Instrument Recognition by XGBoost Combining Feature Fusion. *arXiv:2206.00901 [cs, eess]*. [online]

Available at: https://arxiv.org/abs/2206.00901 [Accessed 25 Oct. 2022].

Li, R. and Zhang, Q. (2022). Audio recognition of Chinese traditional instruments based on machine learning. Cognitive Computation and Systems, 4(2), pp.108–115. doi:10.1049/ccs2.12047.

Miller, J., O'Hanlon, K. and Sandler, M.B. (2022). *Improving Balance in Automatic Chord Recognition with Random Forests*. [online] IEEE Xplore. Available at: https://ieeexplore.ieee.org/abstract/document/9909558 [Accessed 25 Oct. 2022].

Schramm, R., McLeod, A., Steedman, M., Benetos, E. and Conference (ISMIR 2017), 18th I.S. for M.I.R. (2017). Multi-pitch detection and voice assignment for a cappella recordings of multiple singers. *qmro.qmul.ac.uk*. [online] Available at: https://qmro.qmul.ac.uk/xmlui/handle/123456789/25292 [Accessed 25 Oct. 2022]

Musical Pursuits. (2022). *Music Streaming Statistics in 2022 (US & Global Data)*. [online] Available at: https://musicalpursuits.com/music-streaming/#:~:text=Music%20streaming%20makes%20up%2084 [Accessed 27 Oct. 2022].

HeadphonesAddict (2022). *30+ People Listening to Music Statistics & Trends (2022)*. [online] HeadphonesAddict. Available at: https://headphonesaddict.com/listenn